WHAT IS CLAIMED IS:

1. A computer readable medium including instructions readable by a computer which, when implemented, cause the computer to generate synonymous collocations comprising the steps of:

> extracting collocations from a monolingual corpus;
>
> generating candidate synonymous collocations from the extracted collocations; and
>
> selecting synonymous collocations from the candidate synonymous collocations as a function of translation information.

2. The computer readable medium of claim 1 wherein extracting includes parsing the monolingual corpus into dependency triples.

3. The computer readable medium of claim 2 wherein extracting includes calculating weighted mutual information values for the parsed dependency triples.

4. The computer readable medium of claim 1 wherein generating candidate synonymous collocations includes expanding component words of extracted collocations with synonyms of the component words.

5. The computer readable medium of claim 4 wherein expanding component words includes using a thesaurus.

6. The computer readable medium of claim 1 and further comprising translating candidate synonymous collocations to construct a translation set comprising collocation translations for each candidate synonymous collocation.

7. The computer readable medium of claim 6 and further comprising building a language model of probabilities of collocations obtained from a target language corpus.

8. The computer readable medium of claim 6 wherein translating includes using a bilingual lexicon to translate component words of each candidate synonymous collocation to construct the translation set.

9. The computer readable medium of claim 6 wherein selecting comprises selecting synonymous collocations as a function of the collocation translations.

10. The computer readable medium of claim 9 wherein selecting comprises calculating probabilities of the collocation translations.

11. The computer readable medium of claim 6 wherein selecting comprises constructing a translation model of probabilities of candidate synonymous collocations as a function of the collocation translations.

12. The computer readable medium of claim 11 wherein constructing the translation model includes approximating the probabilities as the product of individual translation probabilities of component words of the candidate synonymous collocations and associated collocation translations.

13. The computer readable medium of claim 11 wherein constructing a translation model includes using a bilingual corpus to approximate word translation probabilities of component words of the candidate synonymous collocations.

14. The computer readable medium of claim 6 wherein selecting synonymous collocations comprises generating feature vectors of the candidate synonymous collocations as a function of the translation sets and the translation information.

15. The computer readable medium of claim 14 wherein selecting synonymous collocations comprises calculating a measure of similarity as a function of the feature vectors.

16. The computer readable medium of claim 15 wherein calculating similarity information comprises using the cosine method.

17. The computer readable medium of claim 15 wherein selecting synonymous collocations comprises selecting

synonymous collocations exceeding a similarity threshold.

18. The computer readable medium of claim 1 and further comprising:
parsing input text into at least one collocation;
substituting said at least one collocation in the input text with one of the selected synonymous collocations; and
generating output text using said one of the selected synonymous collocations.

19. A computer readable medium including instruction readable by a computer which, when implemented, cause the computer to generate a sentence comprising the steps of:

   parsing input text into at least one collocation;
   obtaining synonymous collocations selected as a function of translation information; and
   selecting at least one synonymous collocation for said at least one collocation.

20. The computer readable medium of claim 19 wherein obtaining comprises obtaining synonymous collocations from a lexical knowledge base.

21. The computer readable medium of claim 19 wherein selecting comprises selectively substituting synonymous collocations.

22. The computer readable medium of claim 19 wherein selecting comprises automatically substituting synonymous collocations.

23. A method of constructing synonymous collocation information comprising the steps of:

    extracting collocations from unprocessed language corpus;

    generating candidate synonymous collocations from the extracted collocations; and

    selecting synonymous collocations from the candidate synonymous collocations based on translation information.

24. The method of claim 23 wherein the extracted collocations are extracted from unprocessed source language corpus and further comprising:

    parsing the unprocessed source language corpus into dependency triples; and

    recognizing collocations in the parsed source language dependency triples;

    parsing unprocessed target language corpus into dependency triples; and

    recognizing collocations in the parsed source language dependency triples.

25. The method of claim 23 wherein generating candidate synonymous collocations comprises obtaining

synonyms of component words in the extracted collocations.

26. The method of claim 23 and further comprising translating the candidate synonymous collocations into a translation set comprising collocation translations.

27. The method of claim 26 and further comprising calculating translation probabilities of the collocation translations.

28. The method of claim 27 and further comprising generating feature vectors for pairs of candidate synonymous collocations as a function of the collocation translations and associated translation probabilities.

29. The method of claim 28 and further comprising generating similarity values for pairs of feature vectors.

30. The method of claim 29 wherein selecting synonymous collocations comprises selecting pairs of candidate synonymous collocations having similarity values exceeding a selected threshold.

31. A method of generating language comprising the steps of:

    parsing an input sentence into collocations;

accessing a database of synonymous collocations generated using translation information; and

substituting parsed collocations in the input sentence with synonymous collocations from the database.

32. The method of claim 31 wherein parsing includes parsing the input sentence in one language and wherein accessing includes accessing a database of synonymous collocations in a language different from the input language.

33. The method of claim 31 wherein accessing includes accessing synonymous collocations selected as a function of translation information comprising collocation translations and corresponding translation probabilities.

34. The method of claim 33 wherein accessing includes accessing synonymous collocations selected based on similarity values calculated using the collocation translations and corresponding translation probabilities.